Fine-grained Opinion Topic and Polarity Identification

Xiwen Cheng, Feiyu Xu

DFKI
Stuhlsatzenhausweg 3
xiwen.cheng, feiyu.xu@dfki.de

Abstract

This paper presents OMINE, an opinion mining system which aims to identify concepts such as products and their attributes, and analyze their corresponding polarities. Our work pioneers at linking extracted topic terms with domain-specific concepts. Compared with previous work, taking advantage of ontological techniques, OMINE achieves 10% higher recall with the same level precision on the topic extraction task. In addition, making use of opinion patterns for sentiment analysis, OMINE improves the performance of the backup system (NGram) around 6% for positive reviews and 8% for negative ones.

Currently, people who want to get an opinion about a certain product have to go through a large number of product reviews. *Opinion Mining* is a recent discipline at the crossroads of information retrieval, text mining and computational linguistics which tries to detect the opinions expressed in the natural language texts automatically. This paper focuses on two aspects of *Opinion Mining*: topic extraction and polarity analysis. For instance, considering the review

(1) Mileage with the VW Golf is great!

The topic is *mileage with the VW Golf* and the polarity is *positive*. Moreover, the foci in the topic could be more specific and be related to the domain concepts. "*VW Golf*" indicates a car instance and "*mileage*" is a property of a car. The understanding of review (1) is that the property *mileage* of the car *VW Golf* gains a compliment. In this paper, we introduce an implemented system OMINE which aims to identify concepts such as products and their attributes, and analyze their corresponding polarities.

OMINE consists of two modules: 1) ontology-based topic extraction and 2) fine-grained polarity analysis. The former generates a uniform ontology on top of existing domain-specific ones and extends lexicons of the generated ontology to identify concept-related topics. The latter acquires sentiment knowledge (i.e. sentimental lexicons and negations) and generate subjective patterns to train a statistical polarity classifier.

1. Ontology-based Topic Extraction

Even though state-of-the-art approaches have explored many successful strategies to identify topic terms (Liu, et al., 2005; Popescu and Etzioni, 2005; Yi and Niblack, 2005), they did not deal with the work in linking the terms in topics with domain concepts. Considering the example (1), they can not link the term *mileage* in the topic *mileage with the VW Golf* with the concept *a property of a car*. (Popescu and Etzioni, 2005) distinguishes topics as *part* or *property* by means of WordNet IS-A relation and morphological cues (e.g. *-ness*). However, since WordNet is an open-domain tool, its usage is not suitable for domain-specific applications. For instance, there is no way to know that *V-6* and *V-8* are different versions of the automobile engines and to assign *2000 Honda Accord Coupe* as the concept *car entity*. Therefore, our goal at this point is to take advantages of domain-specific ontology to overcome this problem. We do this by performing three steps.

1.1. Offline Ontology Building

Given the set of existing ontologies (i.e. specialized ones w.r.t. car functions, properties and components), OMINE automatically merges two of them gradually until a uniform ontology is generated. In each iteration, it begins with similarity calculation between each pair of concepts from two ontologies. This similarity is a likelihood ratio between the cardinalities of intersection and union of word stems. In another word, it depends on the numbers of same word stems used in two concepts. If there is no shared word stems, the two concepts can not be connected.

Between two ontologies, if the one whose root is connected to any concept of the other, it will be called as the specific ontology; if both of their roots are connected to each other, the one with less concepts is the specific ontology. The remaining ontology is called the general ontology. Finally, the concept in the specific ontology in the button up order will be attached to the connected concept in the general ontology preserving its descendants' hierarchical relations. For each concept in the general ontology, the connected concept in the specific ontology is selected by the constraints in priority order: 1) the concept with the maximum similarity; 2) if more than one concept have the same maximum similarity, we select the one with a maximum depth (i.e. the distance between the concept and the root in the ontology); and 3) if more than one concept have the same maximum depth, we select the first one in the current list.

1.2. Ontology Lexicalization

In order to adapt the generated ontology to real applications, we have to give concepts in the ontology with commonly used expressions (e.g. jargon, abbreviation, and acronym). The straight forward way is to look up each concept in a domain-specific synonymy glossary. However, since it is not easy to obtain this kind of resources, we introduce an algorithm Head-ARGUMENT Matching

(HAM) to handle this case. HAM which is similar to 'head-matching'-heuristic (Cimiano, et al., 2004) aims to retrieve related terms in a glossary for the input concept. This algorithm consists of two parts: 1) it describes each term in the glossary with several dimensions, and 2) it looks up input words in the different combinations of dimensions.
In our case, except of the normal attributes *WORD*, *EXPLANATION*, we add four attributes:

1. *BASICTERM* is included terms, e.g. the term *clutch pedal* has basic terms *clutch* and *pedal*;
2. *EQUATION* is the jargon, abbreviation, acronym which is acquired by ad hoc linguistic patterns: "*term* (*equation term*)" and "*term* known as *equation term*", e.g. *Air conditioning* has abbreviation *A/C*;
3. *COMPONENT* is the included basic term which is also the head of the term, e.g. *brake pedal* is a component *pedal*;
4. *ARGUMENT* is the included basic terms which are not at the head position, e.g. *brake pedal* has an argument *brake*.

The query model combines the dimensions and delivers three solutions. First, if the input concept is one word, HAM looks it up in *BASICTERM* field. Second, if it is a word plus an indicator, e.g. *"system"*, HAM searches the word in *COMPONENT*, or *"part"*, HAM searches the word in *FUNCTION*. Third, if it is a compound word, HAM looks up the head word in *COMPONENT* and the other words in both *FUNCTION* and *EXPLANATION*, e.g. the input is *fuel injector* and we can get {*injector*, *cold start injector*, *saturated switch injector*}). Moreover, the terms in the *EQUATION* field are also saved as lexicons of related concepts.

1.3. IE-based Topic Extraction

For identifying topics with domain concepts, we use a rule-based Name Entity Recognition (NER) and IE engine *SProUT*(Drożdżyński, et al., 2004). All the lexicons in the ontology are utilized to extend SProUT's Typed Feature Structure gazetteer while persevering the ontological concept information (Schäfer, 2006) and we manually compiled typed feature-based rules to identify concept-related topics. For instance, atomic topics like *Passat* can be recognized as the concept *model* by unary patterns *<model=[car_model]>*, and other more complex topics like *Passat TDI 4dr Wagon* can be recognized as a concept *car* by the rule *<car=@seek(model)(property)+>*.

1.4. Experiment

We developed an ontology `CarOnto` on top of the ontologies extracted from online resources, *eBay*[1] and *AutoMSN*[2]. The online automobile glossary *Auto Glossary*[3] is used in the ontology lexicalization step. Our experimental data is 1000 sentences extracted from *User Review* of *AutoMSN*. We manually detected 2038 domain-specific terms as the golden standard.
`CarOnto` obtains 363 concepts (e.g. *Air Intake*), 1233 instances (e.g. *5-speed automatic overdrive*), 145 values of properties (e.g. *wagon* for *Style*, *250@5800 RPM* for *Horsepower*) and 803 makes and models (e.g. *BMW, Z4*). Ontology lexicalization extends 363 concepts to 9033 lexicons. Consequently, `CarOnto` has 11214 domain-related lexicons. `OMINE` achieves the recall of 20.97% and the precision of 88.12% before lexical enrichment, and the recall of 89.35% and the precision of 94.44% after lexical enrichment (see Table 1). Comparatively, TermExtractor[4](Sclano and Velardi, 2007) achieves the best precision of 97.46% while a pretty low recall of 15.72%. Our recall outperforms OPINE(Popescu and Etzioni, 2005) about 17% and the precision about 4%.

	Recall	Precision
TermExtractor	15.72%	97.46%
OPINE	72.13%	90.15%
Before Enrichment	20.97%	88.12%
After Enrichment	**89.35**%	**94.44**%

Table 1: Result of Topic Extraction

1.5. Evaluation

Coverage of term recognition is very high. Consider the following examples:

(2) (a) Best bang for <u>the buck</u>!
(b) A few other <u>acc</u> are wanted.
(c) <u>Handling and riding</u> is good.
(d) <u>The auto trans</u> works very smoothly.

The lexicons, which are extracted from *eBay* and *AutoMSN* and extended with *Auto Glossary*, are able to cover multiple cases: jargon like *the buck* (i.e. *100 miles per hour*); abbreviations like *the auto trans*(i.e. *automatic transmission*); acronyms like *ACC* (i.e. *Automatic Climate Control* or *Active Cruise Control*); terminologies like *power steering pump* and common words which can indicate domain-specific concepts like *handling and riding* (i.e. an indication of the degree of comfort a tire delivers to the passenger).
However, even though `CarOnto` achieves high coverage in current terms of the automobile domain, there are still three kinds of missing cases: 1) new created words, including new makes, models, jargon, etc.; 2) flexible word composition, for instance, freely adding hyphen or space between words (e.g. *GLS-TDI* and *GLS TDI*, *powertrain* and *power train*, *4Dr* and *4-Dr*); 3) spelling checking (e.g. *gas mielage* for *gas mileage*).

IE-based topic extraction achieves high precision at topic extraction and accomplishes the task of concept assignment. The lexicons that come from `CarOnto` consists

[1]http://www.ebay.com
[2]http://autos.msn.com
[3]http://www.autoglossary.com
[4]http://lcl2.di.uniroma1.it/termextractor/

Pattern	Sample
component=(det)?(car_component)+	*dash light, the 1.8 turbo engine*
car=(det)?(@seek(property))?[car_autoentity]	*a hatchback car, this vehicle*
car=(det|@seek(en-year))(car_make|car_model|car_property)	*a SUV, 2000 325i*
car=(@seek(en-year))?(car_make)?[car_model]	*2007 Mazda CX-7)*
car=(@seek(car))(property)+	*2006 Honda Accord Coupe*
car=(@seek(car))(@seek(component))(property)+	*2002 Jetta 1.8T, 2005 VW Passat TDI 4dr Wagon*

Table 2: Typed Feature-based Extraction Pattern

of five basic concepts: *car_make*, *car_model*, *car_property*, *car_component* and *car_autoentity*. Prolific concepts can be recognized by unary patterns: *make=[car_make]*, (e.g. *BMW*); *model=[car_model]*, (e.g. *Jetta*) and *property=[car_property]*, (e.g. *Coupe*). Other more complex expressions are done by patterns in Table 2. The type *make*, *model* and *property* are assigned concepts to the identified terms. Even though, these restrict patterns overcome the problem to identify not only terms but also concepts, they could not deal with the embedded relations yet. Consider the following example:

(3) I love the looks of the interior and exterior.

OMINE assigns the terms *looks*, *interior* and *exterior* as a *Properties* list while fails to specify the correct understanding as *the Properties of the interior* and *the Properties of exterior*. Moreover, all the related works only concern about noun phrases. However, consider the examples below,

(4) (a) This car is stylish.
(b) I get plenty of compliments on how it looks.

the adjective *stylish* in Example (4.a) and the verb *looks* in Example (4.b) gives out the topic *Style* and *Appearance* respectively. These are all interesting topics that we will consider in future.

2. Fine-grained Polarity Analysis

Several researchers have attempted to determine whether a term is a marker of subjective content and what is its sentiment orientation (e.g. positive, negative, neutral) (Hatzivassiloglou and McKeown, 1997; Turney, 2002). However, the sentiment is conveyed not only by single words or phrases but rather by their combinations or contexts. Therefore, some researchers examine whether a given text has a factual nature or expresses an opinion by means of subjective patterns (Riloff and Wiebe, 2003; Riloff, et al., 2006; Popescu and Etzioni, 2005; Wilson, et al., 2005). Taking advantages of above approaches into account, we introduce a two-step learning method: the first step is to acquire sentiment knowledge (i.e. lexical sentiment orientation and negation words), and the second is to train a Naïve Bayes (NB) classifier by subjective patterns which are generated on top of dependency structure with lexical sentiment knowledge.

POSITIVE	NEGATIVE
awesome, cute, speedy excellent, well, standard great, strong, comfortable sporty, super, adorable	unimpressive, awful, terrible useless, tremendous, costly expensive, troublesome, tight cumbersome, ugly, squeaky

Table 4: Sample Result of Sentiment Words

2.1. Acquisition of Sentiment Knowledge

According to the assumption that sentiment words occur frequently in the sentences with corresponding polarities, the task can be solved as relevant term discovery with specific polarities. A term is regarded as a relevant one if it occurs more frequently in a certain category (i.e. *positive*, *negative*, *neutral*) while it occurs occasionally elsewhere. The highest related category is the polarity of a sentiment word. Moreover, considering the observation that negation words always change the polarity of the sentences, we focus on the words which are leaf nodes in the dependency structure (which is acquired by MiniPar (Lin, 2001)) and whose occurrence will always alter the polarity of sentences.

2.2. Polarity Analysis

The subjective pattern is generated on top of a claim. Claim is a simple sentence with at least one topic. If the simple sentence contains sentiment words, claim is the minimum syntactic category containing sentiment words and topics. For each pattern, there are at most three kinds of representations: 1) dependency structure with lexicon information (LOP), 2) if LOP contains sentiment words, these words are replaced by their polarities (SenOP), 3) if LOP or SenOP contains negation words, these words are replaced by the tag "NEG" (NegSenOP). Considering the example in Table 3, each pattern is represented as a *string*. The string transform refers to Penn Treebank[5]. In our case, each element (e.g. *"be":VBE:i*) indicates *STEM:POS:RELATION* (*STEM: Part of Speech: Dependency Relation*).

2.3. Experiment

The corpus is collected from the PROS/CONS part of *User Review* supplied by *AutoMSN*. We simply assume that the sentences in PROS convey positive opinions and the other in CONS are negative ones. We collected around 20 thousand sentences, in which 50% are positive and 50% are negative sentences. To improve the robustness, we use a NB classifier with NGram (Pang, et al., 2002) as our baseline

[5]http://www.cis.upenn.edu/ treebank/

Sentence	*The **Jetta** is the least reliable car.*
LOP(L)	("be":VBE:i)((TOPIC:N:subj)((DET:Det:det))(TOPIC:N:pred)((DET:Det:det)("reliable":ADJ:mod)(("least":ADJ:mod))))
SenOP(S)	("be":VBE:i)((TOPIC:N:subj)((DET:Det:det))(TOPIC:N:pred)((DET:Det:det)(PRO:ADJ:mod)(("least":ADJ:mod))))
NegSenOP(N)	("be":VBE:i)((TOPIC:N:subj)((DET:Det:det))(TOPIC:N:pred)((DET:Det:det)(PRO:ADJ:mod)((NEG:ADJ:mod))))

Table 3: Subjective patterns extracted from the sentence *the Jatte is the least reliable car*

POS	**Negation words**
aux	*doesn't, didn't, wouldn't, shouldn't couldn't, don't, can't, won't*
det	*no, little, least*
mod	*never, barely, not, less*

Table 5: Result of Negation Words

	5000		10000		20000	
X-CV	10-CV		25-CV		50-CV	
N=5	Pro	Con	Pro	Con	Pro	Con
Baseline	86.3		86.3		85.6	
L	86.5	86.6	87.3	87.7	88.2	88.2
L+S	84.9	83.2	83.8	85.8	89.1	87.7
L+S+N	84.8	83.0	84.9	87.6	**91.9**	**93.9**

Table 6: Accuracy of Polarity Analysis

system and the experimental result shows that N=5 makes the classifier achieve the best performance. The final result using multiple patterns with cross validation is given in Table 6 ("L" means we only use LOP as features, "L+S" means we use both LOP and SenOP, and "L+S+N" means we use all of LOP, SenOP and NegSenOP). The experiment performs different fold size cross-validation *X-CV* according to the size of corpus. In the experiment, we used adjectives as candidates and finally acquired 623 sentiment words. Moreover, we got 22 negation words. Among them, 95.0% of sentiment words are correct. The sample results are given in Table 4. On the other hand, for negation words, the precision is 73.8%. The correct ones are given in Table 5. From the result, we observe that the best performance is achieved by using a bigger training corpus with more general representations.

2.3.1. Evaluation

Since the corpus comes from PROS/CONS of *User Review*, it conveys explicit opinions. It helps us to get high precision for sentiment word identification. However, consider the following examples:

(5) (a) The seat of the car is big.

(b) The seat of the car is too big.

If we read Example (5.a), it is impossible to distinguish whether the sentence is subjective or not and what is the polarity conveyed by the adjective *big*. On the other hand, Example (5.b) clearly expresses a negative opinion via *big*. Therefore, we should consider more context information to study this dynamic polarity issue in future.

Most of opinions are conveyed by words directly, e.g, the most frequent pattern (6) (e.g. *a very versatile car*) which conveys positive polarity by the sentiment word *PRO*. However, there are many other ways to express opinions, e.g. subjunctive mood, irony. In the product review, pattern (7), e.g. *I would like more leg room*, *I would like more cup holders* and *I would like more gas efficiency*, gives a negative polarity even though it includes the word *like* which usually occurs in positive utterances. No only these kinds of commonly used patterns, some expressions have attitudes in specific domains. Pattern (8) gives a positive polarity towards TOPIC, which has instances like *the car is loaded with airbags*, *the car is loaded with features*, *this vehicle is loaded with style*.

(6) *<Det "very" PRO TOPIC>*

(7) *<I would like more TOPIC>*

(8) *<TOPIC be loaded with>*

(9) *<TOPIC look expensive>*

(10) *<TOPIC look cheap>*

(11) *<TOPIC look CON>*

The experiment shows the better result with more general patterns. However, there are some exceptions. For instance, the LOP (9) has a positive polarity and another LOP (10) conveys a negative polarity. If these LOPs are generated as a SenOP, *expensive* and *cheap* are both assigned a negative polarity. Therefore, the SenOP (11) leads to errors in polarity assignment. As we known, this issue is also caused by assigning the static and unique polarity to each sentiment word. We plan to acquire dynamic or situational lexical polarities in the future.

3. Conclusion

Compared with state-of-the-art works, `OMINE` succeeded in pioneering implementation of ontological concept assignment to identified topics. It outperforms the baseline system under a large training corpus for polarity analysis, which benefits from the capability of subjective patterns to deal with the sentiment of word combinations. However, there are still a lot of open topics in the real applications. We will intend to overcome them in future.

Acknowledgement

The presented research was partially supported by a grant from the German Federal Ministry of Education and Research to the project Hylap (FKZ: 01IWF02) and EUC-funding for the project RASCALLI. Our special thanks go to the anonymous reviewers for their thorough and highly valuable comments.

4. References

Philipp Cimiano, Aleksander Pivk, Lars Schmidt-Thieme and Steffen Staab. 2004. *Learning Taxonomic Relations from Heterogeneous Evidence*. in *Proceedings of the ECAI 2004 Ontology Learning and Population Workshop*.

Witold Drożdżyński, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. Shallow processing with Unification and Typed Feature Structures - Foundations and Applications. KI, 18(1):17, 2004

D. Lin, LaTaT. 2001 Language and Text Analysis Tools Proc. Human Language Technology Conference, San Diego, California

Bing Liu , Minqing Hu , Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the Web. In *Proceedings of the 14th international conference on World Wide Web*, May 10-14, 2005, Chiba, Japan

Ulrich Schäfer. 2006. *OntoNERdIE—Mapping and Linking Ontologies to Named Entity Recognition and Information Extraction Resources.* In *Proceedings of the 5th International Conference on Language Resources and Evaluation LREC-2006*, Genoa, Italy

Sclano, F. and Velardi, P. TermExtractor: a Web Application to Learn the Shared Terminology of Emergent Web Communities. To appear in *Proc. of the 3rd International Conference on Interoperability for Enterprise Software and Applications (I-ESA 2007).* Funchal (Madeira Island), Portugal, March 28C30th, 2007.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2002)* pp. 79-86, Philadelphia, Pennsylvania.

Ana-Maria Popescu and Oren Etzioni. 2005 Extracting product features and opinions from reviews. In *Proceedings of EMNLP 2005*, pp.339–346.

Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing* p.105-112, July 11, 2003

Ellen Riloff, Siddharth Patwardhan and Janyce Wiebe. Feature Subsumption for Opinion Analysis. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*. pages 440-448, Sydney, Australia, July 2006.

Theresa Wilson, Janyce Wiebe and Paul Hoffmann (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, Vancouver, Canada.

Jeonghee Yi and Wayne Niblack. 2005. Sentiment Mining in WebFountain. In *Proceedings of the 21st International Conference on Data Engineering (ICDE 2005)*, 1084-4627/05.

Vasileios Hatzivassiloglou and Kathy McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97)*, pages 174C181.

Peter D. Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-02)*, pp. 417-424, Philadelphia, Pennsylvania.